CLAIMS

1. A method for controlling a switch comprising:

a number of input ports, each receiving data cells on a respective link;

a number of output ports sharing a buffer space in which each output port can reserve space for an output queue, wherein incoming data cells are switched to an appropriate output queue;

a flow control means for pausing and un-pausing senders on selected links; the method including the steps of:

monitoring the remaining available buffer space AS of the shared buffer;

estimating the expected total content LE of the links;

calculating a free margin (FM) as the remaining available buffer space minus the expected total content of the links FM=AS-LE;

if the free margin sinks below a threshold AS-LE < A, then a selected link is paused;

if the free margin thereafter raises above a threshold AS-LE > B, then a selected paused link is un-paused.

2. A method according to claim 1, wherein the flow control means comprises a pause frame generator for generating pause frames to be sent to data senders in order to pause senders on a selected link, and generating un-pause frames to be sent to data senders in order to un-pause senders on a selected paused link.

3. A method according to claim 2, wherein the content LE of the links is estimated as the sum of the contents of all the input links.

4. A method according to claim 3, wherein the estimation of the content LE of the links takes into account the different link lengths and bit rates.

5. A method according to claim 4, wherein each link estimate is based on a model of the behaviour of each port.

6. A method according to claim 5, wherein the model consists of a curve having different segments, each segment reflecting a specific state of the port.

7. A method according to claim 6, wherein the states include:

a state in which the link is full and contains a maximum amount of data;

a state in which the port is to be paused and is waiting for a pause frame to be sent and in which the link remains at the maximum amount for a fixed duration or until the pause frame is sent;

a state in which a pause frame is sent and the port is waiting for a fixed duration to allow a packet to leave the sender, and in which the link remains at the maximum amount;

a state in which the pause frame has reached the sender and the link content is decreased linearly with time to a minimum amount;

a state in which the port is to be un-paused and is waiting for an un-pause frame to be sent and in which the link remains constant for a fixed duration; and

a state in which an un-pause frame is sent and the link content is increased linearly with time to the maximum amount.

8. A method according to claim 7, wherein the maximum amount of data equals twice as much as a round trip content plus two full-sized packets.

9. A method according to claim 7, wherein the minimum amount of data equals one full-sized packet.

10. A method according to claim 7, wherein the slopes of the linear increase and decrease depend on the bit rate of the respective link.

11. A method according to claim 2, wherein the most offending sender is paused first.

12. A method according to claim 2, wherein the least offending sender is un-paused first.

13. A method according to claim 11 or 12, wherein offending senders are detected by means of an overflow sum counter OFS.

14. A method according to claim 13, wherein a counter OFS is associated with each input port, and is increased each time the input port sends a packet to a congested output port.

15. A method according to claim 14, wherein the counter OFS of each input port is increased with the packet length, each time the input port sends a packet to a congested output port.

16. A method according to claim 14, wherein the counter OFS is reset to zero when its associated input port receives an un-pause frame

17. A method according to claim 14, wherein an output port is considered congested if the queue length thereof exceeds a threshold.

18. A method according to claim 17, wherein the queue length threshold equals a maximum length packet.

19. A method according to claim 13, wherein a maximum value is defined for the OFS counters, and when one counter reaches this maximum, all counters are divided by 2.

20. A method according to claim 13, wherein a maximum value is defined for the OFS counters, and when one counter reaches this maximum, the value of the smallest counter is subtracted from all the counters.

21. A method according to claim 13, wherein all the OFS counters are decreased linearly with time.

22. A method according to claim 1, wherein the threshold A is set to zero (A=0).

23. A method according to claim 1, wherein the threshold A is set to a negative value (A<0).

24. A method according to claim 1, wherein the threshold A is less than or equal to the threshold B ($A \leq B$).

25. A switch comprising:

a number of input ports, each receiving data cells on a respective link;

a number of output ports sharing a buffer space in which each output port can reserve space for an output queue, wherein incoming data cells are switched to an appropriate output queue;

a flow control means for pausing and un-pausing senders on selected links; the switch further including means for:

monitoring the remaining available buffer space AS of the shared buffer;

estimating the expected total content LE of the links;

calculating a free margin (FM) as the remaining available buffer space minus the expected total content of the links FM=AS-LE;

wherein the flow control means is arranged to pause a selected link, if the free margin sinks below a threshold AS-LE < A; and to un-pause a selected paused link, if the free margin thereafter raises above a threshold AS-LE > B.

26. A switch according to claim 25, wherein the flow control means comprises a pause frame generator for generating pause frames to be sent to data senders in order to pause senders on a selected link, and generating un-pause frames to be sent to data senders in order to un-pause senders on a selected paused link.

27. A switch according to claim 26, wherein the content LE of the links is estimated as the sum of the contents of all the input links.

28. A switch according to claim 27, wherein the estimation of the content LE of the links takes into account the different link lengths and bit rates.

29. A switch according to claim 28, wherein each link estimate is based on a model of the behaviour of each port.

30. A switch according to claim 29, wherein the model consists of a curve having different segments, each segment reflecting a specific state of the port.

31. A switch according to claim 30, wherein the states include:

a state in which the link is full and contains a maximum amount of data;

a state in which the port is to be paused and is waiting for a pause frame to be sent and in which the link remains at the maximum amount for a fixed duration or until the pause frame is sent;

a state in which a pause frame is sent and the port is waiting for a fixed duration to allow a packet to leave the sender, and in which the link remains at the maximum amount;

a state in which the pause frame has reached the sender and the link content is decreased linearly with time to a minimum amount;

a state in which the port is to be un-paused and is waiting for an un-pause frame to be sent and in which the link remains constant for a fixed duration; and

a state in which an un-pause frame is sent and the link content is increased linearly with time to the maximum amount.

32. A switch according to claim 31, wherein the maximum amount of data equals twice as much as a round trip content plus two full-sized packets.

33. A switch according to claim 31, wherein the minimum amount of data equals one full-sized packet.

34. A switch according to claim 31, wherein the slopes of the linear increase and decrease depend on the bit rate of the respective link.

35. A switch according to claim 26, wherein the flow control means is arranged to pause the most offending sender first.

36. A switch according to claim 26, wherein the flow control means is arranged to un-pause the least offending sender first.

37. A switch according to claim 35 or 36, wherein the flow control means contain an overflow sum counter OFS to detect offending senders.

38. A switch according to claim 37, wherein a counter OFS is associated with each input port, and is increased each time the input port sends a packet to a congested output port.

39. A switch according to claim 38, wherein the counter OFS of each input port is increased with the packet length, each time the input port sends a packet to a congested output port.

40. A switch according to claim 38, wherein the counter OFS is reset to zero when its associated input port receives an un-pause frame

41. A switch according to claim 38, wherein an output port is considered congested if the queue length thereof exceeds a threshold.

42. A switch according to claim 41, wherein the queue length threshold equals a maximum length packet.

43. A switch according to claim 37, wherein a maximum value is defined for the OFS counters, and when one counter reaches this maximum, all counters are divided by 2.

44. A switch according to claim 37, wherein a maximum value is defined for the OFS counters, and when one counter reaches this maximum, the value of the

smallest counter is subtracted from all the counters.

45. A switch according to claim 37, wherein all the OFS counters are decreased linearly with time.

46. A switch according to claim 25, wherein the threshold A is set to zero (A=0).

47. A switch according to claim 25, wherein the threshold A is set to a negative value (A<0).

48. A switch according to claim 25, wherein the threshold A is less than or equal to the threshold B ($A \leq B$).